NX 奈学教育

大数据集群资源如何评估？

Kafka集群资源评估

企业需求背景

需要构建一个广告流处理平台，需要构建一个Kafka集群，
该集群的目标就是每天要hold住200亿请求

二八法则

QPS估算

每天集群需要承载200亿数据请求，一天24小时，对于网站，晚上12点到凌晨8点这8个小时几乎没多少数据。使用二八法则估计，也就是80%的数据（160亿）会在其余16个小时涌入，而且160亿的80%的数据（128亿）会在这16个小时的20%时间（3小时）涌入。

QPS计算公式=12800000000÷(3*60*60)=118万，故高峰期Kafka集群需要要抗住每秒118万的并发。

存储估算

每天200亿数据，每个请求3kb，也就是55T的数据。如果保存4副本，55*4=220T，保留最近5天的数据。故需要220 * 5 = 1100T

QPS角度

如果资源充足，让高峰期QPS控制在集群能承载的总QPS的30%左右，故目前kafka集群能承载的总QPS为**400万**左右才是安全的，根据经验一台物理机能支持4万QPS是没问题的，所以从QPS的角度讲，需要物理机100台。

磁盘的数量

需要多少个磁盘?

100台物理机，需要存储1100T的数据，每台存储11T的数据，一般的配置是11块盘，一个盘2T就绰绰有余。

log.dirs=/data1,/data2,/data3,....

磁盘的类型

是需要SSD固态硬盘，还是普通SAS机械硬盘？

SSD就是固态硬盘，比机械硬盘要快，SSD的快主要是快在磁盘随机读写

Kafka是顺序写的，机械硬盘顺序写的性能机会跟内存读写的性能是差不多的，所以对于Kafka集群使用机械硬盘就可以了。

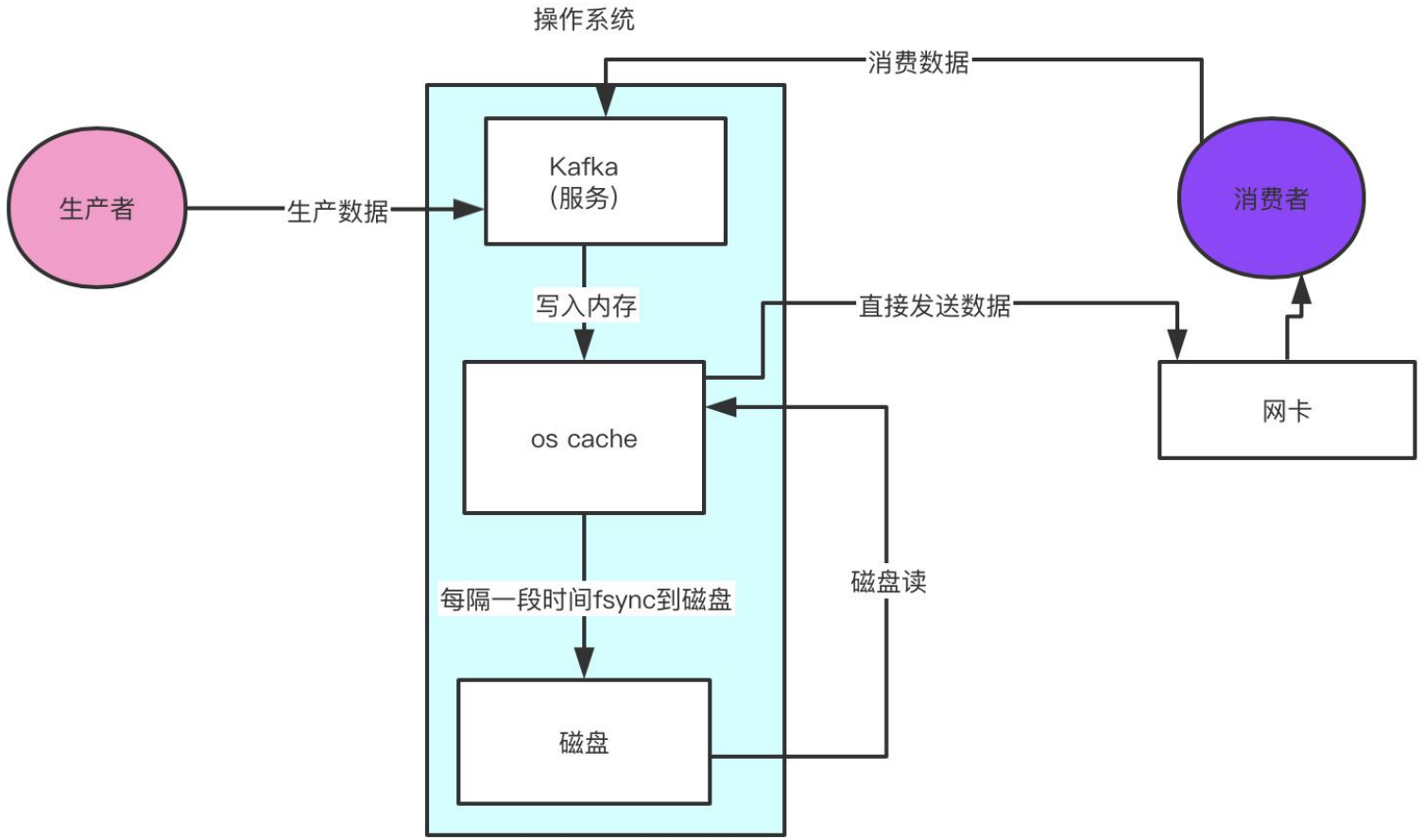
内存角度

Kafka自身的jvm是用不了过多堆内存，因为kafka设计就是规避掉用jvm对象来保存数据，避免频繁fullgc导致的问题，所以一般kafka自身的jvm堆内存，分配个6G左右就够了，剩下的内存全部留给os cache。

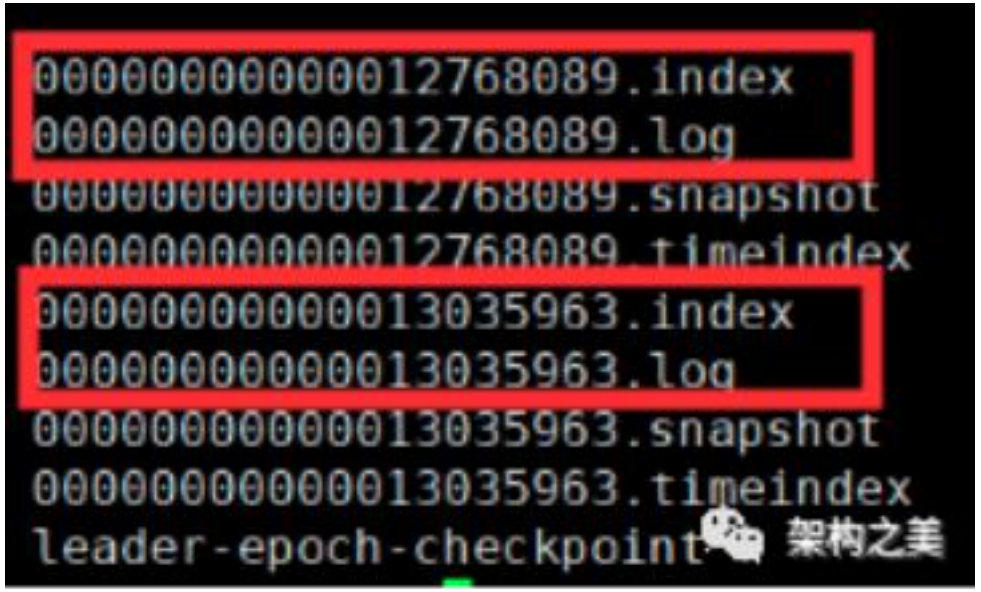
每台服务器多要多少内存呢?

内存角度

Kafka读写流程回顾



内存角度



经过梳理，此集群有3个topic，这3个topic的partition的数据在os cache里效果是最好的。3个topic，一个topic有500个partition。那么总共会有1500个partition。每个partition的Log文件大小是1G，我们有4个副本，也就是说要把1500个topic的partition数据都驻留在内存里需要6000G的内存。我们现在有100台服务器，所以平均下来每天服务器需要60G的内存，但是其实partition的数据我们没必要所有的都要驻留在内存里面，50%的数据在内存就非常好了（25%的数据在内存也可以），60G * 0.5 = 30G就可以了。所以一共需要36G的内存，故我们可以挑选64G内存的服务器也非常够用了，

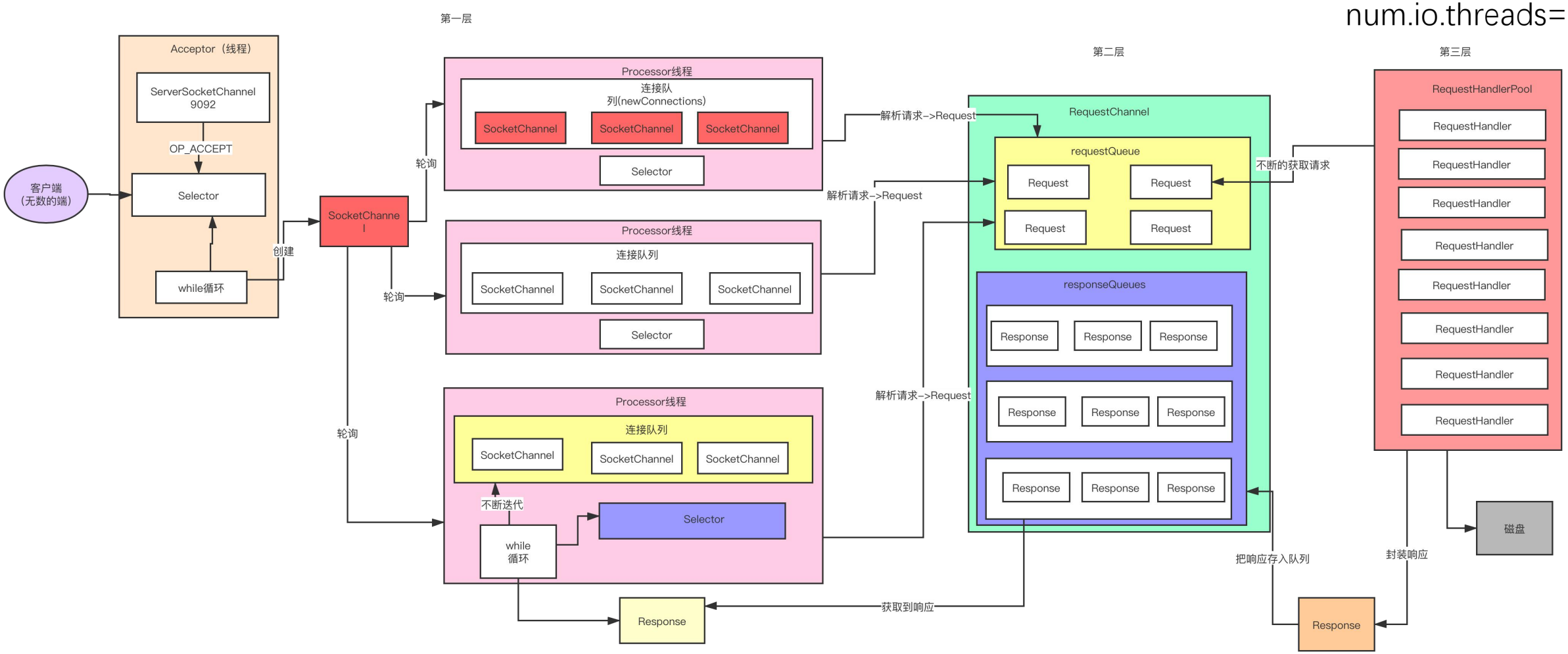
CPU角度

CPU规划，主要是看Kafka进程里会有多少个线程，线程主要是依托多核CPU来执行的，如果线程特别多，但是CPU核很少，就会导致CPU负载很高，会导致整体工作线程执行的效率不太高。

那Kafka进程里面大概有多少线程呢?

CPU角度

num.io.threads=8



num.network.threads=3

CPU角度

1. Accept线程
2. 默认的3个Process线程（一般会设置程9个）
3. 默认的8RequestHandle线程（可以设置成32个）
4. 清理日志的线程
5. 感知Controller状态的线程
6. 副本同步的线程

……

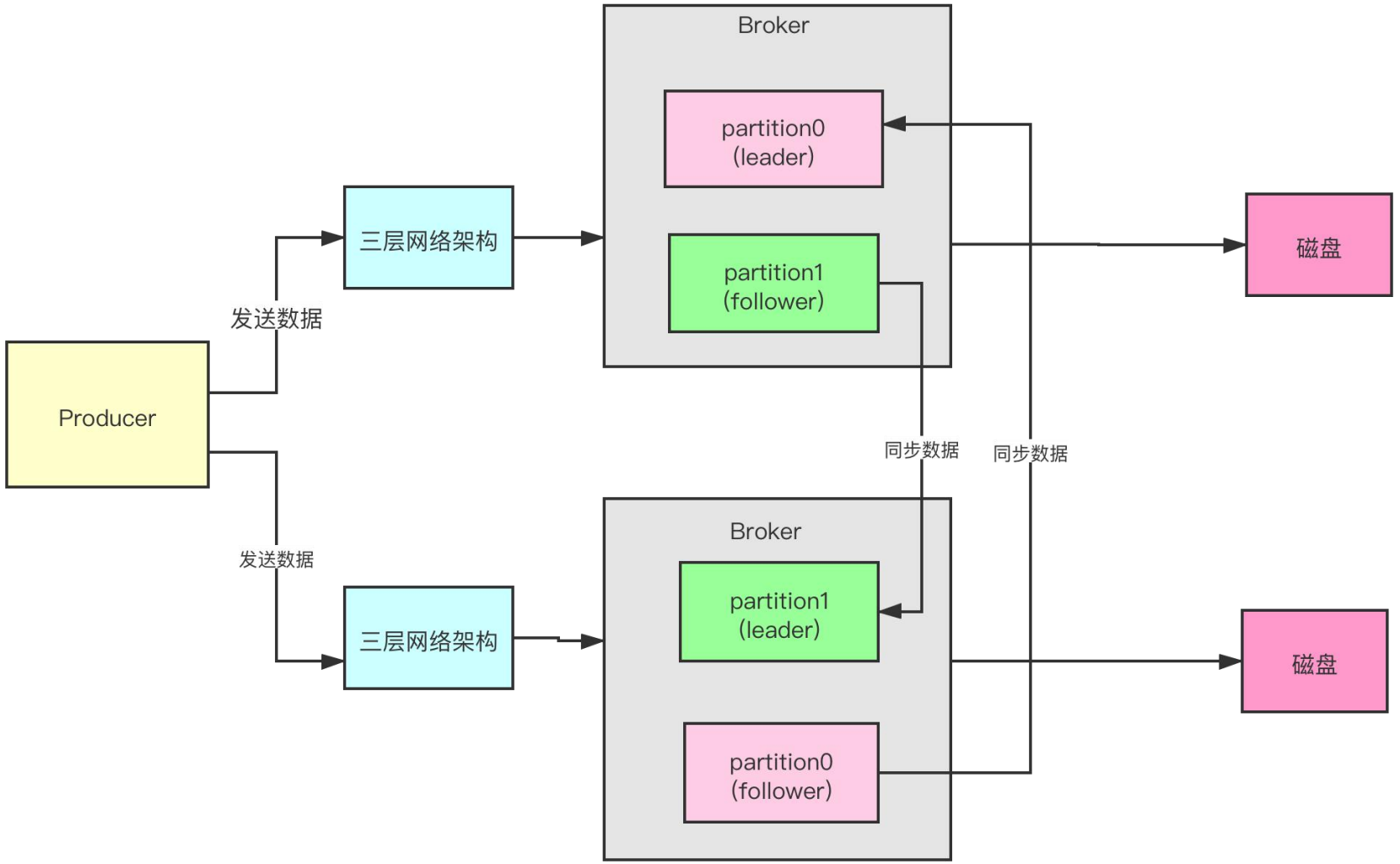
估算下来Kafka内部有100多个线程

CPU角度

4个cpu core，一般来说几十个线程，在高峰期CPU几乎都快打满了。**8个cpu core**，也就能够比较宽裕的支撑**几十个线程**繁忙的工作。所以Kafka的服务器一般是建议16核，基本上可以hold住一两百线程的工作。当然如果可以给到32 cpu core那就最好不过了！

网卡角度

1. 接收请求
2. 副本同步
3. 消费数据



网卡角度

每秒两台broker机器之间大概会传输多大的数据量？高峰期每秒大概会涌入118万个请求，约每台处理1.18万个请求，每个请求3kb，故每秒约进来34M数据，我们还有副本同步数据，故高峰期的时候需要34M * 4 = 136M/s的网络带宽，34 * 10（consumer）=340M/s，故总的需要340M+136M=476M，所以在高峰期的时候，使用千兆网卡即可

集群资源规划

200亿请求，118w/s的吞吐量，1100T的数据，100台物理机

硬盘：11（SAS） * 2T，7200转

内存：64GB，JVM分配6G，剩余的给os cache

CPU：16核/32核

网络：千兆网卡

企业需求背景

奈学教育

需要构建一个广告流处理平台，该平台离线数据需要永久存储

大数据集群资源评估

每天55T，**副本数为3**，预估一年的存储资源。

一年需要的存储资源：55 * 3 * 365 = 1095T

数据需要**进行加工**（建模）：1095 * 5 = 5475 T

数据增速是**每个季度40%**，5475 * （1.4 ^ 4）= 21032T

磁盘只能存到**80%**，故需要26290T的存储空间

机器配置：32cpu core, 128G内存，11 * 7T

故：21032/77 = **341台**服务器

需要构建一个广告流处理平台，实时任务需要的计算资源

实时集群资源规划

cpu : 实时任务需要的计算资源，因为我们有1500个partition，故需要1500 cpu core

内存： cpu和内存根据经验值是1:4 故需要6000G资源

机器配置： 内存 128G　cpu core　32

从cpu角度，1500/32 = 需要 46台服务器
从内存角度，6000/128 =需要 46台服务器

redis资源规划

redis需要抗的并发和kafka是一样的（存储不是瓶颈），故Redis集群在安全的情况下，需要抗住**400万QPS**，按照每台redis 抗**10万请求**来算，需要**40台**服务器。